

Claims

What is claimed is:

1. A system for parsing XML, the system comprising:
 - a scanner adapted to parse an XML stream to locate at least one XML token associated with an XML item;
 - a reader adapted to selectively pull the XML item from the XML stream;
 - and
 - a retriever adapted to retrieve information associated with the pulled XML item.
2. The system of claim 1 where the XML item is one of a start token, an end token, markup, content, an entity reference, an external reference, an element, a tag, character data, an attribute, a CDATA section, a comment and a processing instruction.
3. The system of claim 1 further comprising a checker adapted to determine whether the pulled XML item is well-formed.
4. The system of claim 1 further comprising a validator adapted to determine whether the pulled XML item is valid.
5. The system of claim 1, where the scanner is further adapted to facilitate navigating a virtual node in a stream of XML nodes.
6. The system of claim 5, where the scanner is further adapted to resolve an external reference in the XML stream.
7. The system of claim 6, where the reader is further adapted to selectively pull an XML node from the stream of XML nodes.

8. The system of claim 7, where the reader is further adapted to selectively pull an XML node from the stream of XML nodes based, at least in part, on data provided to the reader by a parse requestor.
9. The system of claim 1, where the retriever is further adapted to expose data model and/or Infoset information associated with the pulled XML item.
10. The system of claim 3, where the checker is further adapted to determine whether the pulled XML item is well-formed based, at least in part, on comparing the pulled XML item to one or more syntax documents.
11. The system of claim 4, where the validator is further adapted to determine whether the pulled XML item is valid based, at least on part, on comparing the XML item to one or more DTD, schema, and external data representation documents.
12. The system of claim 1, where at least one of the scanner, the reader and the retriever are an object.
13. The system of claim 3, where the checker is an object.
14. The system of claim 4, where the validator is an object.
15. A computer readable medium storing computer executable components of a system for parsing XML, the system comprising:
 - a component for locating an XML token that identifies at least one XML item;
 - a component for selectively pulling the XML item;
 - a component for retrieving metadata associated with the pulled XML item;
 - a component for verifying whether the pulled XML item is well-formed;
 - and

a component for verifying whether the pulled XML item is valid.

16. A method for parsing XML, the method comprising:
 - instantiating a pull model parser;
 - establishing a state associated with the pull model parser;
 - accepting a parse request;
 - selectively pulling an XML item based, at least in part, on the parse request; and
 - updating the state based on the selectively pulled XML item.
17. The method of claim 16 further comprising checking the pulled XML item to determine whether it is well-formed.
18. The method of claim 17 where determining whether the pulled XML item is well-formed comprises comparing the pulled XML item to one or more syntax documents.
19. The method of claim 16 further comprising checking the pulled XML item to determine whether it is valid.
20. The method of claim 19 where determining whether the pulled XML item is valid comprises comparing the pulled XML item to at least one of a DTD, a schema and an external data representation.
21. The method of claim 17 further comprising checking the pulled XML item to determine whether it is valid.
22. The method of claim 16 where instantiating the pull model parser comprises:
 - associating a stream with the pull model parser; and
 - initializing a scanner adapted to facilitate navigating within the stream.

23. The method of claim 16 where establishing the state associated with the pull model parser comprises:
 - associating a state machine with the pull model parser; and
 - establishing an initial state position within the state machine.
24. The method of claim 16 where selectively pulling an XML item further comprises:
 - positioning a virtual node over an XML node within a stream of input XML nodes; and
 - selectively extracting an XML item from the XML node over which the virtual node is positioned.
25. The method of claim 24 where selectively pulling an XML item further comprises resolving an external reference in the XML item.
26. The method of claim 16 where updating the state based on the selectively pulled XML item comprises repositioning the state position within the state machine.
27. A computer readable medium storing computer executable instructions for a method for parsing XML, the method comprising:
 - operably connecting a pull model parser and a state machine;
 - establishing an initial state in the state machine;
 - accepting a parse request;
 - selectively pulling an XML item identified in the parse request;
 - maintaining the state machine in response to one or more events associated with parsing and/or pulling the pulled XML item;
 - checking the pulled XML item to determine whether it is well-formed; and
 - checking the pulled XML item to determine whether it is valid.

28. A system for parsing XML, the system comprising:
 - means for tracking a state associated with the pull model parser;
 - means for abstracting input and/or output associated with the pull model parser;
 - means for accepting one or more parse requests;
 - means for selectively pulling an XML item based, at least in part, on the one or more parse requests;
 - means for determining whether the pulled XML is well-formed; and
 - means for determining whether the pulled XML item is valid.
29. A data packet adapted to be transmitted between two or more computer processes, the data packet comprising:
 - one or more first fields adapted to store a pulled XML item; and
 - one or more second fields adapted to store metadata information associated with the pulled XML item.
30. A system for parsing XML, the system comprising:
 - an extractor adapted to pull an XML item from an input source; and
 - a parser adapted to selectively parse the XML item into one or more sub-items.
31. The system of claim 30, where the sub-items comprise at least one of a start token, an end token, markup, content, an entity reference, an external reference, an element, a tag, character data, an attribute, a CDATA section, a comment and a processing instruction.
32. The system of claim 30, further comprising a metadata retriever adapted to retrieve information associated with the pulled XML item.

33. The system of claim 32 where the information associated with the pulled XML item comprises at least one of a namespace name, a local name, a prefix, an unordered set of attributes, and an ordered list of children.
34. The system of claim 32 further comprising a syntax analyzer adapted to determine whether the pulled XML item conforms to a desired XML syntax.
35. The system of claim 34 further comprising a semantic analyzer adapted to determine whether the pulled XML item conforms to a desired XML semantic.
36. The system of claim 35 further comprising a style analyzer adapted to determine whether the pulled XML item conforms to a desired style sheet.
37. A method for parsing XML, the method comprising:
pulling an XML item from an input source; and
selectively parsing the XML item into one or more sub-items.
38. The method of claim 37, where sub-items comprise at least one of a start token, an end token, markup, content, an entity reference, an external reference, an element, a tag, character data, an attribute, a CDATA section, a comment and a processing instruction.
39. The method of claim 38, further comprising retrieving metadata information associated with the pulled XML item.
40. The method of claim 39 where the metadata information associated with the pulled XML item comprises at least one of a namespace name, a local name, a prefix, an unordered set of attributes, and an ordered list of children.

41. The method of claim 39 further comprising analyzing syntax associated with the pulled XML item to determine whether the syntax conforms to a desired XML syntax.
42. The method of claim 41 further comprising analyzing semantics associated with the pulled XML item to determine whether the semantics conform to a desired XML semantics.
43. The method of claim 42 further comprising analyzing the style of the pulled XML item to determine whether the pulled XML item conforms to a selected style sheet.